基于文献计量的国内数据素养研究热点分析

魏建松

南昌师范学院 南昌 330000

摘要：[目的/意义]数据素养已然成为生活在数字时代的个体必备的基本素养。[方法/过程]对已公开发表的数据素养主题文献进行梳理有助于明晰数据素养相关研究脉络，进一步优化后续相关研究。通过搜集、整理并清洗CNKI收录的SCI、EI、CSSCI、CSCD和北大核心来源期刊中发表的学术论文展开文献计量分析，系统地梳理了数据素养研究脉络。[结果/结论]目前国内学者在数据素养的内涵与价值研究、数据素养能力评价和数据素养培养路径方面进行了初步探索，借助CiteSpace分析发现，“数据驱动”是数据素养相关研究中的最新突现词，这可能是当前学界在数据素养研究领域中的研究前沿。

关键词：数据素养；知识图谱；文献计量

引言

长期以来，以直觉和个人阅历作为主要支撑的决策过程由于随意性强、缺乏客观依据等问题而饱受诟病。随着各行业数字基础设施的持续建设，文字、图像等各类信息载体在以数字编码形式进行存储和传输等方面难度逐渐减低，而大数据等信息技术的蓬勃发展，计算技术与分析技术的突破，又促使蕴含着碎片化信息的社会经济、现实世界、管理决策等片段的大数据得以聚合、分析以及应用，通过解读一系列的碎片化信息，进一步强化了数据的生产要素与生产力功能，深刻改变了人类的传统思维方式和生产生活方式，数据化发展也逐渐成为当前社会信息化发展的重要趋势。

伴随着新一代信息技术驱动各领域的深度变革，各行各业的信息化环境也在发生着前所未有的变化，各类数据的获取、存储以及开发等能力成为行业从业人员的基本能力。数据素养作为大数据环境下合理获取、理解、运用、评价和管理数据的综合能力，是信息素养、和统计素养等概念的延伸，是个体在信息化环境中创新应用各类数据开展社会生产生活等实践活动的必备素养。

为此，本文利用文献计量研究方法，对国内公开发表的重要学术期刊中与数据素养主题相关的学术文献开展统计分析、主题挖掘及文献对比分析，了解数据素养的相关研究在国内的发展趋势。为国内学

---

*本文系江西省高等学校教学改革研究项目“基于增值评价的教学质量评价改革研究”（项目编号：JXJG-22-23-30）、南昌师范学院教育评估项目“教师数据素养能力评价与提升路径研究”（项目编号：21XJZX30）研究成果之一。

作者简介： 魏建松，教务处质量科副科长，讲师，硕士研究生，E-mail：jason@ncnu.edu.cn

者捕捉研究动态和研究议题提供决策参考。

1 研究方法与数据来源

近年来，文献资料的爆炸式增长以及文献计量研究工具的不断优化，使得深入文献计量研究领域的学者日益增多。文献计量学是通过数学与统计学方法的结合，对已有文献数据收集整理并进行分析、评估和预测的一种方法，在文献主题、学者、机构和期刊等方面能够从多个维度实现较为全面的识别。

1.1 研究方法

在文献计量法的运用中，大部分学者侧重于从如下几个方面进行研究：1）以特定学科领域内的核心期刊为数据源，运用文献计量法梳理该学科研究脉络，识别当前学科热点，预测学科研究方向。例如：丁秀莲等学者基于管理科学与工程领域内的 46 份国际权威期刊上发表的论文数据，运用文献计量法梳理了该学科领域内近 10 年的研究热点[1]。2）运用文献计量法进行学术期刊评价研究。俞立平等学者在 z 指数原理的基础上提出采用低被引论文比的倒数代替被引集中度，进而得到一种评价学术期刊的改进版复合文献计量指标[2]。除此之外，还有学者运用文献计量法研究不同地区学者对某一领域的贡献[3]。

1.2 数据来源

为了提升数据素养研究的文献数据质量，同时避免国内外数据库重合的文献数据对本研究造成干扰，文献来源限定为中国知网中来源类别为 SCI、EI、CSSCI、CSCD、北大核心和 AMI 中收录的学术论文。

根据文献检索中常用的“主题”“篇名”“关键词”“摘要”以及“全文”检索等方法的特征，综合权衡文献数据查全率和查准率，本研究选择“主题”进行检索。在国内中文期刊数据库中国知网中使用高级检索功能，通过设定查询时间范围为 2022 年及以前，主题词设定为“数据素养”，来源类别限定为 SCI、EI、CSSCI、CSCD 和北大核心，初步得到文献数据 565 条。

为提升研究结果的准确性，在进行数据分析前，通过数据清洗进一步提高文献查准率，剔除述评、会议报道、成果介绍、卷首语、无作者以及其他与本研究无关条目，最终得到有效文献 502 篇。

2 研究趋势的基本特征

2.1 发文量年度增长趋势

通常情况下，科学知识的增长量与该主题文献发文量的变化有着紧密联系，学界在衡量科研产出与热点时，往往将某主题文献的发文量作为一项重要指标，文献发文量的高低在一定程度上反映出该主题

的研究热度和成熟度。

通过梳理和对比最终得到的文献发表情况可以发现，以“数据素养”为主题的文献首次出现于2010年，图 1 “数据素养”主题文献发文量年度趋势中显示，文献发文量从2015年左右开始攀升，但在2016至2019年间发文量增长趋势趋于平缓。

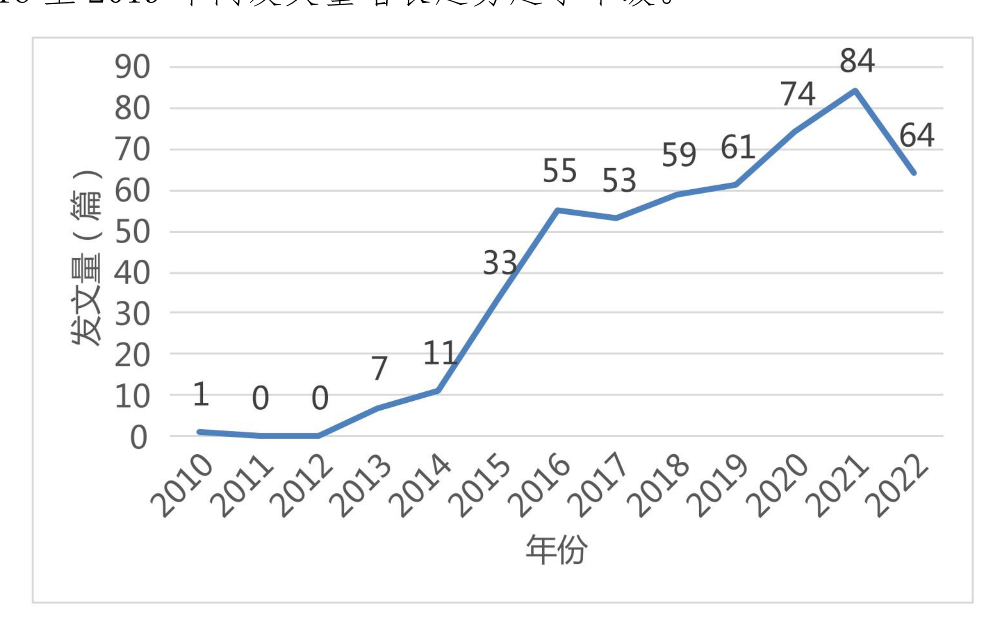


图 1 “数据素养”主题文献发文量年度趋势

2.2 发文期刊分布分析

根据信息管理理论，分析信息分布情况，有助于认识信息分布的集中、分散特征和规律。侯剑华（2015）等学者通过对我国期刊评价的研究，发现发表期刊和研究方向中也存在着马太效应[4]。在对已发表的文献发文期刊分布情况进行统计后发现，共有130家期刊发表过数据素养研究相关论文，其中发文量排名前10的来源期刊刊发论文共计206篇，占比41.4%；发文量超过10篇的来源期刊刊发论文共计251篇，占比50%，具体来源期刊见表 1，通过进一步分析可知，在以“数据素养”为主题的相关研究中，约10%的期刊发表的论文数量占全部文献数量的50%以上，体现出“数据素养”研究论文在少数期刊中的富集现象。这些发文量排名靠前的期刊均为图书情报和教育教学研究领域的期刊。

表 1 数据素养研究发文量超过10篇的来源期刊

发文量排名	期刊名称	篇数	影响因子
1	图书与情报	32	3.264
2	图书馆学研究	29	1.907
3	图书情报工作	26	2.632
4	图书馆理论与实践	24	1.109

5	图书馆工作与研究	18	2.751
6	情报理论与实践	18	2.623
7	电化教育研究	17	4.503
8	图书馆	16	2.167
9	情报科学	13	2.43
10	现代情报	13	2.435
11	新世纪图书馆	13	0.77
12	现代教育技术	12	2.329
13	数字图书馆论坛	10	1.573
14	情报资料工作	10	2.764

注：表中影响因子数据来源于中国知网 2022 版综合影响因子。

2.3 作者合作网络分析

同一学者在同一主题下发文量的数量越多，往往代表该学者在该领域的研究更深入，也体现出学者在该领域可能有着引领作用，其学术成果也可能是该研究领域思想和观点的重要渠道。利用 CiteSpace 对作者合作网络进行分析时，数据筛选中选择 g-index，在未简化网络的情况下得到节点数 269 个，连线 139 条，在一定程度上表明学界关于数据素养的研究还较为分散，在对作者合作网络可视化分析时将阈值设定为 3 时得到图 2 作者合作网络，图中可以看出作者合作网络较为稀疏，从发文量统计来看，仅有杨现民、黄如花、胡卉及邓李君等学者在“数据素养”研究领域发文量较多，其中杨现民聚焦教育教学领域中的数据素养研究，主要以教师数据素养研究为主，而黄如花、胡卉及邓李君则主要聚焦图书馆学领域中的数据素养研究。从图 2 中也可以看出这几位发文量高的学者还未形成合作网络。

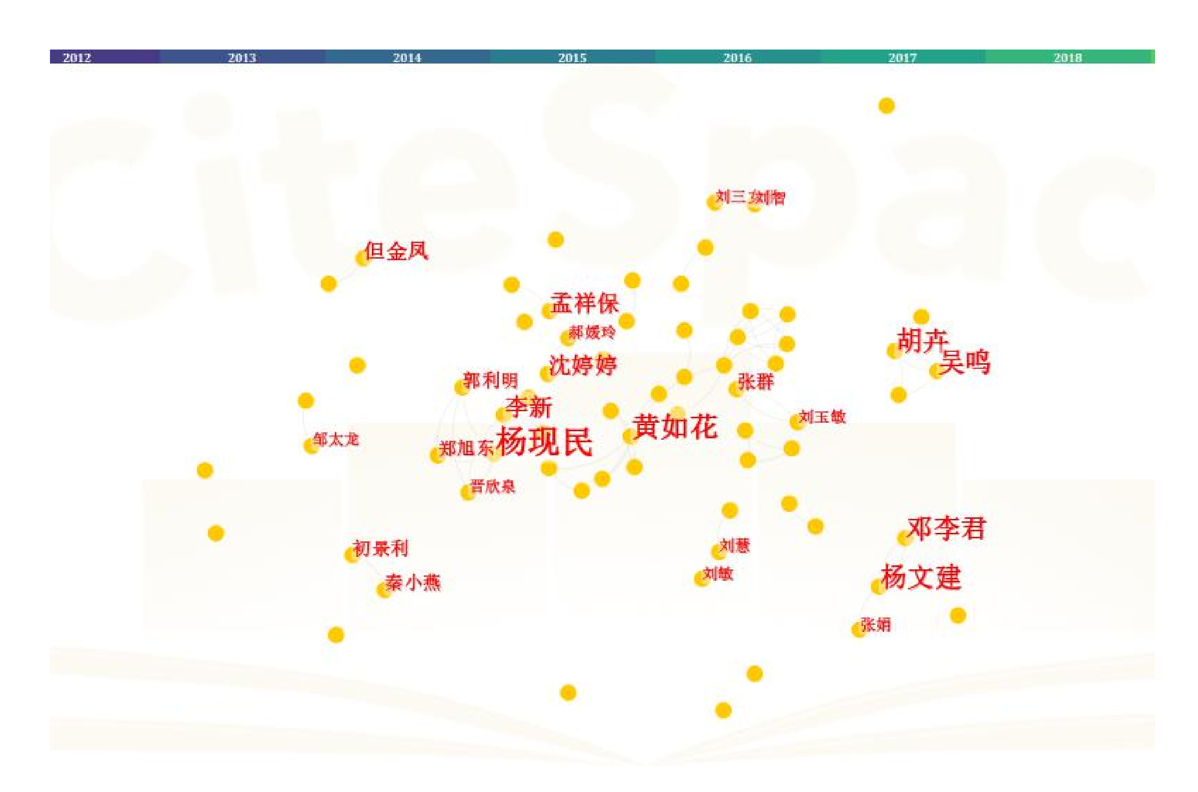


图 2 作者合作网络

2.4 研究机构及合作网络分析

同样借助 CiteSpace 软件，依然在数据筛选中选择 g-index，并将阈值设定为 5 时得到研究机构合作网络如图 3 所示，可以看出国内各研究机构之间在该领域内的合作关系并不强，但存在一些合作，发文量前五的研究机构分别为武汉大学信息管理学院、中国科学院文献情报中心、中国科学院大学、中山大学资讯管理学院以及四川外国语大学图书馆。

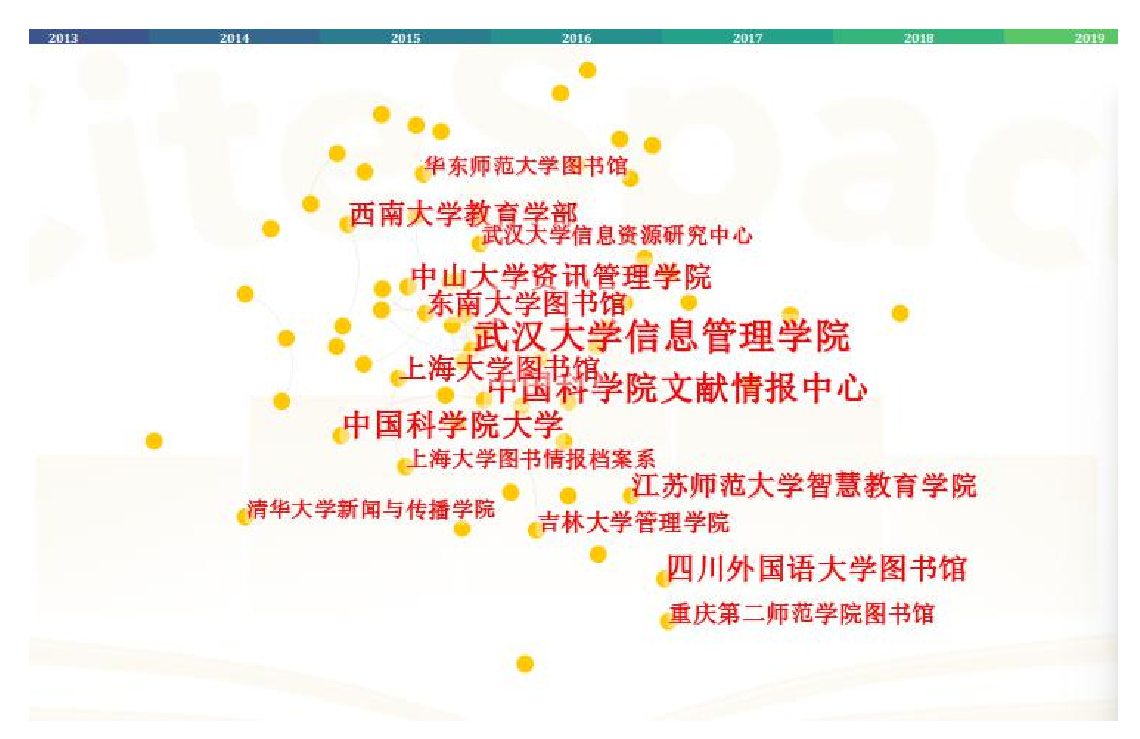


图 3 研究结构合作网络

3.基于关键词研究的知识图谱

通过对"数据素养"主题研究的文章关键词共现知识图谱的分析，能够了解数据素养研究领域的知识聚类和演化规律，并获得该研究主题的共引轨迹和演进网络，从而有助于提炼出当前学界关于"数据素养"主题的研究热点并明晰演变趋势。

3.1 基于关键词共现网络的研究热点分析

文献关键词是对研究内容的高度概括，在 CiteSpcae 中通过节点大小可以判断出关键词出现的频率高低，高频词代表着当前研究领域的研究热点，各节点之间的连线表示其代表的关键词出现在同一文献中。在未简化网络的情况下，根据 CiteSpace 计算得到节点 323 个，连线 575 条，将阈值设定为 24 并对关键词进行聚类分析，调整显示效果后得到图 4 所示关键词共现及聚类图谱。

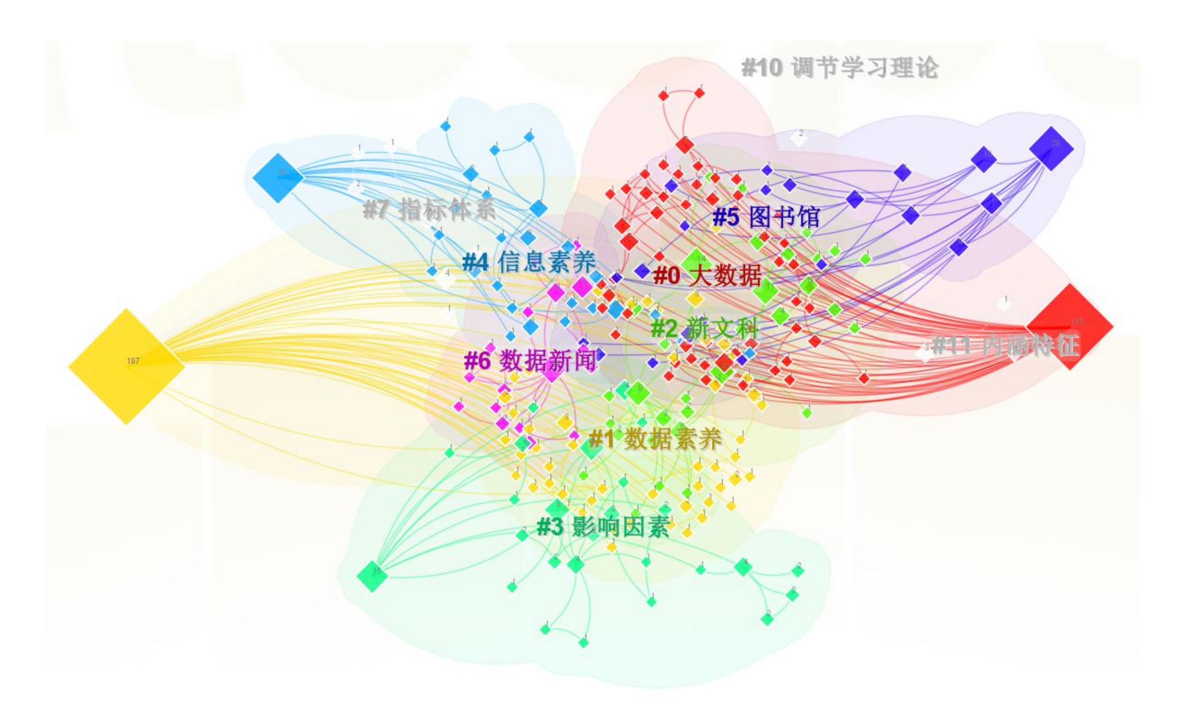


图 4 数据素养研究关键词共现及聚类图谱

频次前十的关键词为数据素养、大数据、信息素养、图书馆、科学数据、人工智能、数据驱动、数据新闻、数据馆员、数据服务，各关键词频次及中心性如表 2 所示。

表 2 2010-2022 年数据素养研究高频关键词统计

序号	频次	中心性	关键词	序号	频次	中心性	关键词
1	197	0.83	数据素养	6	14	0.03	人工智能
2	113	0.43	大数据	7	11	0.03	数据驱动
3	38	0.07	信息素养	8	10	0.01	数据新闻
4	29	0.03	图书馆	9	10	0.02	数据馆员
5	15	0.08	科学数据	10	9	0.02	数据服务

3.2 基于关键词的聚类分析

在对关键词进行聚类分析后得到聚类模块值 Q 值为 0.5819，聚类平均轮廓值 S 值为 0.907，分别大于 0.3 和 0.7，表明本文聚类结构合理有效。在 CiteSpace 中，默认情况下，少于 10 篇文献的聚类不显示，因此图 4 中只显示 10 组聚类标签，依次为：#0 大数据研究；#1 数据素养研究；#2 新文科研究；#3 影响因素研究；#4 信息素养研究；#5 图书馆研究；#6 数据新闻研究；#7 指标体系研究；#10 调节学习理论研究；#11 内涵特征研究。将图 4 转换成 Timeline 视图后得到图 5。

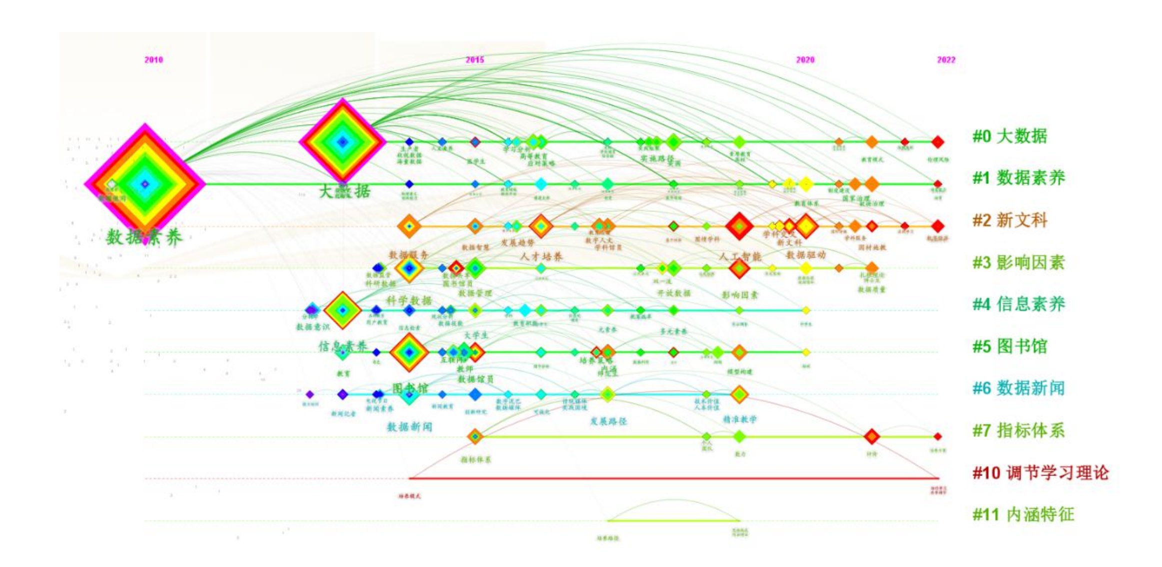


图 5 2010-2022 年数据素养研究关键词共现 Timeline 视图

基于关键词聚类并进一步分析已发表的数据素养相关研究论文后可将研究热点提炼为:

（1）数据素养内涵与价值研究

关于数据素养内涵与价值研究这一主题,学界主要围绕教学科研人员和图书馆员进行研究。数据素养是图书馆员和科研工作者的核心素养，也是数据管理服务研究领域的一个关键概念[5]。对教师而言，具有数据素养的教师能够较好地适应基于数据密集型的科学研究模式，在科研成果产出等方面也有促进作用[6]，同时，也有助于教师专业发展从“粗糙的经验”向“基于证据”的发展范式转变[7]。对学生而言，不同学段学生的数据素养构成维度有所差异[8-9]，但善于利用大数据获取所需信息，能够帮助自己提升专业知识储备，并更好地为自己步入社会奠定知识基础[9]。

（2）数据素养能力评价研究

随着学界对数据素养内涵研究的不断深入，关于数据素养能力的评价也逐渐增多，研究对象主要集中在高校师生之中。数据素养是在大数据环境下社会对个体素养的新要求，是对传统信息素养的继承和发展[10]。在高校教师、博士生、硕士生以及本科生群体中，不同群体在数据素养能力方面具有较大差异[11]。而在评价指标体系研究中，学者们从不同维度开展了一系列研究。如：李青、赵欢欢（2018）采用多种研究方法，归纳和分析筛选出的数据素养构成要素相关文献，从数据知识、数据技能、教学应用以及意识道德四个方面构建了教师数据素养评价指标体系[12]。而马腾、孙玲（2019）基于信息生态理论构从信息、信息环境和信息技术等维度构建出评价指标体系，从而实现对学生数据素养评价[13]。在中小学学生中，还可以从数据知识和技能、

数据思维、数据意识和数据道德规范等层面构建能力模型[14]。通过结合高中数学的教学特点，还有学者从 4 个维度设计了数据分析素养的评价框架[15]。经过对已发表的数据素养评价相关的研究可以发现，在目前已有的能力评价研究中，大多数评价指标体系的构建均包含数据意识、数据处理能力、数据交流和数据评价等维度。

（3）数据素养培养路径研究

要使不同领域的数据产生价值，就需要培养不同领域从业者的数据素养能力。对比哈佛大学的教育实践，我国各高校的数据管理服务处于无序状态，在大数据时代，高校应先制定数据素养教育和管理政策[16]。对教师（师范生）这一主体而言，数据素养的培养所涉及的知识面广，是一个系统工程，需要多方共同努力[17]。针对中小学教师，可借鉴调节学习理论，构建基于调节支架的中小学教师数据素养培养教学模式[18]。针对科研人员，基于数据生命周期与科研项目生命周期为理论基础的数据素养培养框架的提出，在一定程度上能够让科研人员摆脱数据素养能力不足的困境[19]。

3.3 基于突现分析方法的前沿分析

利用突现分析方法可以对某一研究领域中的关键节点进行深入分析，进而找出当前活跃或前沿主题。由于中国知网导出的数据中不包含参考文献数据，因此仅对关键词进行突现分析。在保持默认参数不变的情况下，经过计算，得到两个突现关键词，分别为“数据新闻”（强度为 3.3）和“数据驱动”（强度为 3.71），“数据新闻”关键词突现时间为 2014-2017 年。而“数据驱动”关键词突现时间为 2020 年-2022 年。该关键词在突现时间段内的来源文献具体如表 3 所示。

表 3 关键词“数据驱动”的来源文献

序号	作者	发表年份	期刊名称	篇名
1	何亚丽；赵庆香；肖鹏	2020	图书馆论坛	数据驱动时代的图书馆战略规划及其实施策略
2	顾佳妮；杨现民；郑旭东；郭利明	2020	现代远程教育研究	数据驱动学校治理现代化的逻辑框架与实践探索
3	杨现民；郭利明；王东丽；邢蓓蓓	2020	现代远程教育研究	数据驱动教育治理现代化：实践框架、现实挑战与实施路径
4	王世伟	2020	情报资料工作	数据驱动的图书馆学情报学——2019 年图书馆学情报学热点鸟瞰

5	陈曦	2020	教育理论与实践	高校数据驱动教学改革的发展路径
6	赵晏强；周伯柱	2021	图书馆学研究	学科馆员 3.0 及其服务体系构建
7	林攀登；张立国；周釜宇	2021	当代教育科学	从经验回顾到数据驱动：人工智能赋能教师教学反思新样态
8	钟亚平；吴彰忠；陈小平	2021	体育科学	数据驱动精准训练：理论内涵、实现框架与推进路径
9	袁丽；张金华	2021	当代教育论坛	数据驱动赋能新时代教师领导力的思考
10	魏亚丽；张亮	2022	当代教育科学	从“基于经验”到“数据驱动”：大数据时代的教学新样态
11	张黎；赵磊磊	2022	教育学术月刊	中小学教师数据素养的影响因素与提升路径——基于混合研究方法的实证分析
12	杨淑婷；魏非	2022	现代教育技术	教师数据驱动决策能力评价的国际研究述评：框架、工具与实施建议
13	杨现民；张瑶	2022	中国远程教育	教育规模化与个性化矛盾何以破解？——数据驱动规模化因材施教的逻辑框架与实践路径
14	张学波；林书兵；朱莹	2022	中国电化教育	从心理干预到能力构建：教师数据素养培育的系统框架与实践机制
15	胡艺龄；张启迪；孙珂；吴忭	2022	中国远程教育	中小学教师数据素养培养模式及其应用研究

根据突现分析方法得到的突现词“数据驱动”分析可知，该词的来源期刊依然主要以教育教学和图书馆学研究为主题，又因该词突现时间段为 2020 年-2022 年，因此可将“数据驱动”视为当前数据素养研究前沿。

3.3.1“数据驱动”在图书馆学情报学领域的研究

进入 21 世纪以来，大数据持续成为学术界研究的热点，不同的学科领域，关于大数据的研究维度也在逐渐细分，在对数据驱动的图书馆学情报学研究中，“数据”作为发展基点，强有力的驱动着图书馆学情报学的发展[20]。然而，要将数据驱动落实为图书馆未来的发展方向仍需宏观层面的努力，何亚丽（2020）等学者通过调研美国国家医学图书馆、大英图书馆等国外 15 所图书馆后认为，在面对数据驱

动转型的过程中，应该让数据成为图书馆决策管理的支撑点、业务重构的发力点以及用户服务的增长点[21]。在微观层面，新型科研范式的兴起促使以数据驱动为主要特征、以大数据技术应用为依托、以多源数据为服务基础的新一代学科馆员随之产生，其功能主要体现在数据支持科技决策、科研管理和科研过程的知识服务中[22]。

3.3.2“数据驱动”在教育教学领域的研究

在教育教学领域，数据驱动带来的治理变革对加速实现学校治理现代化具有重要作用[23]。相较于其他传统行业，我国高等教育领域创新力度不足，教育教学和人才培养模式存在批量化生产的弊端，这也导致国内高校在人才培养方面存在同质化问题严重的情况，因此，高校应该将数据作为依托，遵循学习个性化、管理精细化和教学信息化等价值取向，探索利用大数据实现高校教育教学方式与形态变革的有效方法[24]。通过构建学校、政府以及社会三者之间的新型关系，能够促使教育治理业务与数据之间的全面深度融合，通过发挥各主体的积极性提升教育治理水平[25]。近年来，“精准”一词使用频率大幅提高，应用范围也越来越广。在运动训练领域，“精准训练”和“数据驱动”已成为高频词和主旋律，数据驱动精准训练的优势能够让运动员根据个体差异获得最宜刺激，进而产生最优适应，提升训练质量[26]。而如果教师数据素养不高则导致数据驱动赋能教师领导力面临着因材施教不够便捷、不够精细等诸多问题[27]。但要实现“数据驱动”型教学的有效运行，教师需要清晰地认识到教学数据的价值体现应该始终围绕学生这一主体，将冰冷的数据应用于温暖的教学中[28]。

4 结束语

通过对中国知网收录的 SCI、EI、CSSCI、CSCD 和北大核心来源期刊中发表的与数据素养相关的研究文献进行多角度展开分析，得到以下几个结论：

第一，数据素养相关的研究已经成为图书馆学、情报学、教育学重要期刊中持续增长的热点议题之一。大数据、信息素养以及科学数据应用等研究在图书馆学、情报学、教育学期刊中已经占有重要的一席之地，而且相关研究正从数据管理向数据驱动迈进。

第二，从明确提及数据素养的文献中观察到，早期的数据素养研究主要探讨大数据时代数据素养的内涵与价值，这类研究推动了后续研究的发展。大量的数据素养构成、评价与教育培养相关的研究日渐丰富多样。

第三，在已公开发表的研究文献中，大部分研究主要集中于学生、教师以及科研人员群体。大数据时代，数字经济已然成为经济增长的

新动力，对于生活在数字化时代的人们而言，数据素养将是21世纪促进人们发展的必备素养之一。

本文的研究不可避免地存在着局限性。首先，在研究文献的数据来源方面，虽然选取的是中国知网收录的SCI、EI、CSSCI、CSCD和北大核心来源期刊中发表的学术研究论文，具有一定的代表性，但整体覆盖面还不够广。其次，国外数据素养相关研究未纳入分析范围，缺乏对国外相关研究热点和前沿研究主题。

参考文献

[1]丁秀莲,吴强,张朋,等. 基于管理类国际权威期刊的中国管理科学与工程学科研究现状分析[J]. 管理学报，2022，19(2)：159-168.

[2]俞立平,王作功. z指数评价学术期刊的适用性及其改进研究[J]. 情报学报，2018，37(11)：1132-1139.

[3]何佳讯,葛佳烨,张凡. 中国学者管理学研究的世界贡献：国际合作、前沿热点与贡献路径——基于世界千种管理学英文期刊论文(2013～2019年)的定量分析[J]. 管理世界,2021,37(9)：36-67.

[4]侯剑华,刘波. 期刊评价研究产出中的马太效应[J]. 中国科技期刊研究，2015，26(9)：992-996.

[5]孟祥保,常娥,叶兰. 数据素养研究:源起、现状与展望[J]. 中国图书馆学报，2016，42(2)：109-126.

[6]张进良,李保臻. 大数据背景下教师数据素养的内涵、价值与发展路径[J]. 电化教育研究，2015，36(7)：14-19，34.

[7]赵虹元. 基于数据素养的中小学教师专业发展:内涵与路径[J]. 继续教育研究，2017，(10)：77-80.

[8]贾璞,宋乃庆. 大数据时代中学生数据素养：内涵、价值与构成维度[J]. 电化教育研究，2020，41(12)：28-34，58.

[9]张明海,周艳红. 大数据时代大学生数据素养教育的目标定位及体系构建[J]. 图书馆，2016，(10)：84-88.

[10]邓李君,杨文建. 个体数据素养评价体系及相关指标内涵研究进展[J]. 图书情报工作，2017，61(3)：140-147.

[11]隆茜. 数据素养能力指标体系构建及高校师生数据素养能力现状调查与分析[J]. 图书馆，2015，(12)：51-56，62.

[12]李青,赵欢欢. 教师数据素养评价指标体系研究[J]. 电化教育研究，2018，39(10)：104-110.

[13]马腾,孙玲. 信息生态视域下高校大学生数据素养评价研究[J]. 情报科学，2019，37(8)：120-126.

[14]惠恭健,曾磊. 智能时代的数据素养:模型构建、指标体系与培养路径——基于国内外模型的比较分析[J]. 远程教育杂志, 2021, 39(4): 52-61.
[15]陈建明,孙小军,杨博谛. 数据分析素养的评价框架与实施路径研究[J]. 数学教育学报, 2022, 31(2): 8-12, 57.
[16]郝媛玲,沈婷婷,高珊. 高校数据素养教育实践的思考和建议——基于哈佛大学案例和我国图书情报人员访谈的分析[J]. 图书情报工作, 2015, 59(12): 44-51.
[17]张斌,刘三女牙,刘智,等. 基于大数据的师范生数据素养培养策略研究[J]. 电化教育研究, 2017, 38(12): 86-91, 120.
[18]胡艺龄,张启迪,孙珂,等. 中小学教师数据素养培养模式及其应用研究[J]. 中国远程教育, 2022, (3): 51-60.
[19]张军. 面向科研第四范式的科研人员数据素养培养研究[J]. 图书与情报, 2016, (2): 133-136.
[20]王世伟. 数据驱动的图书馆学情报学——2019 年图书馆学情报学热点鸟瞰[J]. 情报资料工作, 2020, 41(1): 39-44.
[21]何亚丽,赵庆香,肖鹏. 数据驱动时代的图书馆战略规划及其实施策略[J]. 图书馆论坛, 2020, 40(11): 98-104.
[22]赵晏强,周伯柱. 学科馆员 3.0 及其服务体系构建[J]. 图书馆学研究, 2021, (14): 79-86.
[23]顾佳妮,杨现民,郑旭东,等. 数据驱动学校治理现代化的逻辑框架与实践探索[J]. 现代远程教育研究, 2020, 32(5): 25-34.
[24]陈曦. 高校数据驱动教学改革的发展路径[J]. 教育理论与实践, 2020, 40(33): 52-55.
[25]杨现民,郭利明,王东丽,等. 数据驱动教育治理现代化: 实践框架、现实挑战与实施路径[J]. 现代远程教育研究, 2020, 32(2): 73-84.
[26]钟亚平,吴彰忠,陈小平. 数据驱动精准训练: 理论内涵、实现框架与推进路径[J]. 体育科学, 2021, 41(12): 48-61.
[27]袁丽,张金华. 数据驱动赋能新时代教师领导力的思考[J]. 当代教育论坛, 2021, (1): 82-88.
[28] 魏亚丽,张亮.从“基于经验”到“数据驱动”：大数据时代的教学新样态[J].当代教育科学,2022,(02):50-56.